

# Principal Component Analysis for the Classification of Cardiac Motion Abnormalities Based on Echocardiographic Strain and Strain Rate Imaging

Mahdi Tabassian<sup>1,2</sup>, Martino Alessandrini<sup>2</sup>, Luca De Marchi<sup>1</sup>, Guido Masetti<sup>1</sup>, Nicholas Cauwenberghs<sup>3</sup>, Tatiana Kouznetsova<sup>3</sup> and Jan D'hooge<sup>2</sup>

<sup>1</sup> Dep. of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy

`mahdi.tabassian2@unibo.it`

<sup>2</sup> Dep. of Cardiovascular Sciences, Cardiovascular Imaging and Dynamics Group, KU Leuven, Leuven, Belgium

<sup>3</sup> Dep. of Cardiovascular Sciences, Research Unit of Hypertension and Cardiovascular Epidemiology, KU Leuven, Leuven, Belgium

**Abstract.** Clinical value of the quantitative assessment of regional myocardial function through segmental strain and strain rate has already been demonstrated. Traditional methods for diagnosing heart diseases are based on values extracted at specific time points during the cardiac cycle, known as ‘techno-markers’, and as a consequence they may fail to provide an appropriate description of the strain (rate) characteristics. This study concerns the statistical analysis of the whole cardiac cycle by the Principal Component Analysis (PCA) method and modeling the major patterns of the strain (rate) curves. Experimental outcomes show that the PCA features can outperform their traditional counterparts in categorizing healthy and infarcted myocardial segments and are able to drive considerable benefit to a classification system by properly modeling the complex structure of the strain rate traces.

**Keywords:** strain/strain rate classification, Principal Component Analysis, feature extraction

## 1 Introduction

Echocardiography is the modality of choice in clinical diagnostics and for the noninvasive assessment of heart function. In daily clinical practice, visual evaluation is widely used to determine regional abnormalities in myocardial wall motion. Although this qualitative assessment can be done easily, it suffers from the inter-observer variability which reduces its clinical value.

Tissue Doppler imaging and speckle tracking are two promising echocardiographic techniques that have been developed for the noninvasive study of myocardial function. Based on these techniques, strain (rate) imaging has been

introduced to provide an effective approach for the assessment of changes in the regional myocardial wall motion deformation [6]. End-systolic strain and peak-systolic strain rate values are two traditional techno-markers that have been extensively used by clinicians to describe the strain (rate) profiles and to classify different heart diseases [9]. These traditional features, however, ignore the diastolic period of the cardiac cycle. They also represent the value of the strain (rate) profile at only one time point and as a result, cannot capture the temporal information available in the deformation curves.

Despite several studies for the classification of regional myocardial function based on the traditional features of the strain (rate) curves [7], [9], only a few investigations have been carried out for the detection of heart abnormalities by taking into account the whole temporal behavior of the strain (rate) curves. The idea of analyzing the whole strain profiles, derived from tagged magnetic resonance imaging, by PCA was initially proposed in [3]. The authors showed that the statistical reference model that achieved by employing PCA and normal strain curves can properly detect abnormal strain patterns. Inspired by [3], PCA has been used in [1] to model ultrasonic strain and strain rate traces of the healthy subjects and it has been discussed that the PCA attributes can provide more information about strain (rate) curves than the traditional features. In [10], the artificial neural network (ANN) was used to classify strain profiles obtained at baseline and during experimentally induced acute ischemia using animal data. In a pre-processing step, each strain curve was represented by 70 equidistant samples and normalized in amplitude. The obtained profiles were then given to an ANN for categorization.

Following the results presented in [1], this paper addresses the subsequent issues: (i) building two PCA models by making use of the normal and acute infarcted strain and strain rate curves, (ii) incorporating the PCA and traditional features in a classification system and examining their capabilities for the categorization of normal and acute infarcted strain and strain rate traces and, (iii) comparing the PCA features extracted from the strain and strain rate traces in terms of the amount of the discriminatory information that they provide for a classification system.

The rest of the paper is organized as follows. Section 2 presents the details of the data acquisition procedure, reviews the PCA method and introduces the employed classifiers. In Section 3, the outcomes of the PCA implementation and the classification phase are presented. Discussion about the obtained results are given in Section 4. Finally, Section 5 draws conclusions and summarizes the paper.

## 2 Material and Methods

### 2.1 Data Acquisition and Preprocessing

A group of 27 normal subjects and 54 subjects with acute myocardial infarction was used in this study. For the patients, myocardial segments were categorized

into infarct, border and remote based on MRI-delayed enhancement and the perfusion territory of the infarct-related vessel [9]. Data acquisition was performed at high frame rate ( $> 180s^{-1}$ ) with a GE VingMed Vivid7 equipped with a 2.5 MHz transducer. For each subject, data were acquired in the apical 2-,3- and 4-chamber views with optimization of the pulse repetition frequency in order to avoid aliasing. An event-driven graphical user interface called SPEQLE [4] was used for the post-processing of the data to extract longitudinal strain (rate) traces in an 18-segment model of the left ventricle [2]. Since the number of samples of the extracted curves could be different due to the differences in the heart rates of the subjects, a linear interpolation procedure was adopted to have the same number of samples in all traces. To avoid unwanted changes of the curves due to the interpolation procedure, each of the six cardiac phases (i.e. electromechanical coupling, isovolumetric contraction, ejection, isovolumetric relaxation, early filling and late filling [6]) was interpolated separately and then merged to have the whole heart cycle. The interpolated curves were then used in the PCA implementation and classification phases.

Table 1 lists the number of subjects that were selected randomly from the healthy and pathological groups for building training, validation and test sets. This random selection was repeated 10 times and the results presented in Section 3 are the average of running the classifier on these 10 different sets of data. Since for a pathological subject only the subset of acutely infarcted segments was used for the classification task, the number of utilized pathological subjects in Table 1 was more than the healthy ones so that both groups had roughly the same number of curves in the training, validation and test sets. Note that, the segmental strain and strain rate curves were sorted in two different groups of training, validation and test data to study their clinical relevance for discriminating normal and infarcted traces.

**Table 1.** Number of subjects taken from the healthy and pathological groups for the training, validation and test sets

	#Training	#Validation	#Test
Healthy Subjects	12	5	10
Pathological Subjects	25	9	20

## 2.2 Principal Component Analysis

PCA is a popular statistical approach for feature extraction, dimensionality reduction and data visualization [8], [12]. Given a data set of random vectors  $X = \{x_1 \dots x_N\}$  where  $x_i \in \mathbb{R}^n$ , the PCA algorithm gives a representation of the data in  $\mathbb{R}^m$  ( $m < n$ ) such that the new variables are less redundant compared to the original ones.

The first step in building the PCA model is *centering* the data that can be done by first computing the mean vector of  $X$  and then subtracting every data vector from it. In the second step, the covariance matrix of the centralized data set is computed. The third step is to find the  $n \times n$  eigenvector matrix  $\Phi = \{\phi_1 \dots \phi_n\}$  and diagonal eigenvalue matrix  $\Lambda = \text{diag}\{\lambda_1 \dots \lambda_n\}$  of the covariance matrix. The final step is to insert the first  $m$  eigenvectors of  $\Phi$  with the largest eigenvalues, which are known as Principal Components (PCs), into a new matrix  $\Psi$  and project the data onto the space spanned by the eigenvectors of  $\Psi$ ,

$$Y = \Psi^T \overline{X} \quad (1)$$

where the variance of the new low-dimensional data  $Y$  is maximized which means that the first axis has the largest variance, the second axis has the second largest variance and so on.

### 2.3 Classifier

Performance of a classification system depends on both the features extracted from the data and the classification technique. In order to differentiate between the effects of the employed features and the classification strategy on the final classification outcomes, two different classifiers are used in our experiments. The first classifier is called the *locan-mean based* (LMB) method [11] and considers the information of classes around a test sample for its classification. The second classifier is *support vector machine* (SVM) [5]. It works based on the idea of the maximum margin solution and finds a hyper-plane that has the greatest distance to the training samples in the boundaries of a binary classification problem.

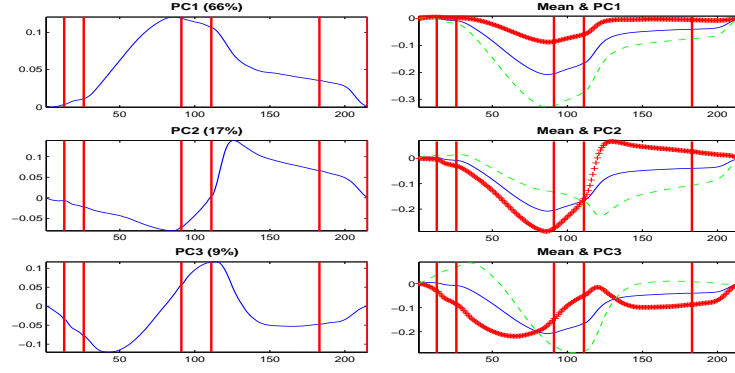
## 3 Results

### 3.1 PCA Outcomes

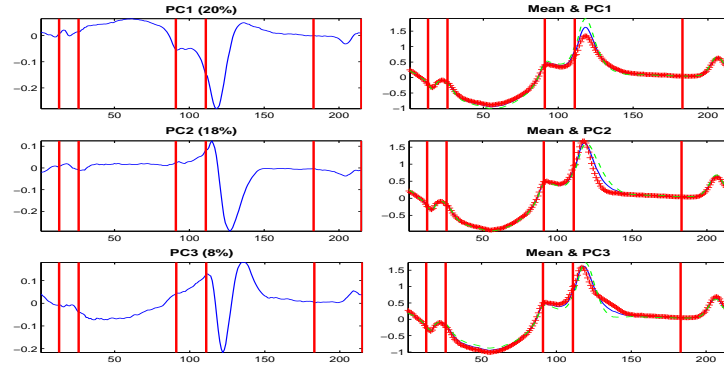
The segmental strain and strain rate traces of the training sets were used to build two separate PCA models. Figs. 1 and 2 illustrate the first three PCs of the strain and strain rate traces, respectively. The variance percentages accounted for the PCs and the results of adding and subtracting the PCs to and from the mean curves are also showed. In order to investigate the PCs' structures in terms of timing of the six mechanical phases of the cardiac cycle, the timing of the onset of each phase is shown with a red vertical line.

### 3.2 Classification Outcomes

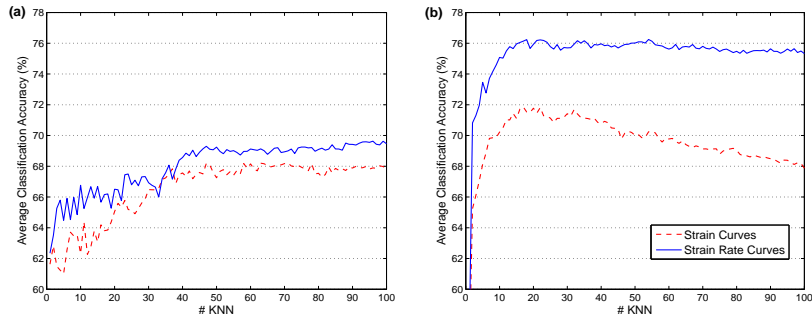
Average classification results on the test data for the LMB and SVM are shown in Fig. 3 and Table 3, respectively. Table 2 gives the best average classification accuracies and their corresponding sensitivity and specificity values obtained by the LMB and selecting K nearest neighbors (KNNs) from [1 100] interval. The number of PCs was set to 10 based on the favorable results gained with the validation data.



**Fig. 1.** The first three PCs of the strain curves, their contributions in the covariance matrix and the results of adding (red ++ ) and subtracting (green - - ) them to and from the mean strain profile



**Fig. 2.** The first three PCs of the strain rate curves, their contributions in the covariance matrix and the results of adding (red ++ ) and subtracting (green - - ) them to and from the mean strain rate profile



**Fig. 3.** Average classification accuracies (%) on the test strain and strain rate curves with the LMB method using (a) the traditional features and, (b) the PCA features

## 4 Discussion

PCA can be considered as a linear model for approximating a set of data in which the main structures of the data are captured by the PCs and their eigenvalues [12]. Figs. 1 and 2 show that for both the strain and strain rate curves, PC1s and the mean curves are very similar in timing and shape except that they have opposite signs which is trivial in linear modeling and can be compensated by multiplying the weights of those PCs by -1. This similarity implies that the first PCs, which have noticeable contributions in data modeling, capture the average patterns of the strain and strain rate curves. The second and third PCs of the strain and strain rate curves, however, reflect important variations during early diastole and model more complicated variations around the mean. As an instance, while the first PC of the strain traces follows the mean curve, the second and third PCs yield to one and two intersections with the mean curve, respectively.

**Table 2.** Best average classification accuracies (%) and their corresponding sensitivity and specificity values of the LMB method with the traditional and PCA features on the test strain and strain rate curves

	Accuracy	Sensitivity	Specificity
<b>Traditional Features</b>			
Strain	68.22	67.51	68.86
Strain Rate	69.63	73.35	66.35
<b>PCA Features</b>			
Strain	71.86	64.98	78.90
Strain Rate	76.26	69.93	83.44

**Table 3.** Average classification accuracies (%) and their sensitivity and specificity values of the SVM method with the traditional and PCA features on the test strain and strain rate curves

	Accuracy	Sensitivity	Specificity
<b>Traditional Features</b>			
Strain	69.61	55.14	83.39
Strain Rate	71.17	67.23	75.56
<b>PCA Features</b>			
Strain	70.06	73.41	67.07
Strain Rate	74.32	77.03	72.28

A key observation that can be made by comparing the percentages of the PCs belonging to the strain and strain rate curves is that in the first case, 92 % of the variances accounted for the whole eigenvectors are captured by the first three PCs while in the latter case this percentage is 46 %. This indicates that the strain rate traces have more diverse and complex temporal behavior than the strain curves.

Tables 2 and 3 show that regardless of the classification methodology, the PCA features yielded to more satisfactory accuracy rates for both the strain and strain rate traces compared to the traditional features. This means that by exploring the whole cardiac cycle instead of only the systolic period, considerably more information about the characteristics of the strain and strain rate curves can be obtained. Another key point that is highlighted in Tables 2 and 3 is the difference between the classification accuracies obtained with features extracted from the strain and strain rate traces. While by using the PCA features, the classification accuracies on the strain rate curves are remarkably better than those of the strain traces, there is not much benefit to be gained by employing the traditional features and the accuracies obtained for the strain rate traces are slightly better than the strain curves. This observation suggests that the strain rate traces provide more discriminatory information than the strain curves and these extra information could be captured by an efficient statistical model like PCA.

Tables 2 and 3 also show that for both the PCA and traditional features, the sensitivity and specificity rates achieved by the LMB and SVM varied considerably. It means that the classification strategy has a direct effect on the final outcomes and suggests that different classification methodologies should be examined to find a suitable setup. These results also imply that by combining several classifiers that address the classification problem from different points of view, a good compromise between the sensitivity and specificity values could be obtained that is a topic for the future research.

## 5 Conclusion

In this study, it was hypothesized that the temporal behavior of the segmental strain (rate) curves contains valuable diagnostic information which can be captured by a rigorous statistical approach. The PCA method was then employed to statistically analyze the strain (rate) traces. In order to evaluate the usefulness of the PCA features, they were compared with the end-systolic strain and peak-systolic strain rate values as the traditional features. Experiments with a data set of strain (rate) curves of healthy and pathological subjects demonstrated that the PCs can provide more discriminatory information for the classification system than the traditional features. Our experiments also showed that analyzing the strain rate traces with PCA would lead to better results than the strain curves. However, more thorough analysis with a larger set of data is needed to improve the classification performance and to determine the role of different parameters that affect the obtained outcomes.

## References

1. Aoued, F., Eroglu, E., Herbots, L., Rademakers, F., D'hooge, J.: A statistical model-based approach for the detection of abnormal cardiac deformation. In: Ultrasonics Symposium, IEEE, vol. 1, 512 - 515 (2005)
2. Cerqueira, M. D., Weissman, N. J., Dilsizian, V., Jacobs, A. K., Kaul, S., Laskey, W. K., Pennell, D. J., Rumberger, J., Ryan, T., Verani, M. S.: Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation*, vol. 105, 539-542 (2002).
3. Clarysse, P., Han, M., Croisille, P., Magnin, I.: Exploratory analysis of the spatio-temporal deformation of the myocardium during systole from tagged mri. *IEEE Transactions on Biomedical Engineering*, vol. no. 11, 1328 - 1339 (2002)
4. Claus, P., D'hooge, J., Langeland, T. M., Bijnens, B., Sutherland, G. R.: SPEQLE (Software Package for Echocardiographic Quantification LEuven) an Integrated Approach to Ultrasound-Based Cardiac Deformation Quantification. In: *Computers in Cardiology*, IEEE, vol. 29, 69 - 72 (2002)
5. Cristianini, N., Shawe-Taylor, J.: *An Introduction to Support Vector Machines*. Cambridge University Press, Cambridge (2000).
6. D'hooge, J., Bijnens, B., Thoen, J., Van de Werf, F., Sutherland, G., Suetens, P.: Echocardiographic strain and strain-rate imaging: a new tool to study regional myocardial function. *IEEE Transactions on Medical Imaging*, vol. 21, no. 9, 1022 - 1030 (2002)
7. Jamal, F., Kukulski, T., Sutherland, G. R., Weidemann, F., D'hooge, J., Bijnens, B., Derumeaux, G.: Can Changes in Systolic Longitudinal Deformation Quantify Regional Myocardial Function After an Acute Infarction? An Ultrasonic Strain Rate and Strain Study. *Journal of the American Society of Echocardiography*, vol. 15, no. 7, 723-730 (2002)
8. Jolliffe, I. T.: *Principal Component Analysis* (Second ed.), Springer (2002)
9. Herbots, L., D'hooge, J., Eroglu, E., Thijs, D., Ganame, J., Claus, P., Dubois, C., Theunissen, K., Bogaert, J., Dens, J., Kalantzi, M., Dymarkowski, S., Bijnens, B., Belmans, A., Boogaerts, M., Sutherland, G., Van de Werf, F., Rademakers, F., Janssens, S.: Improved regional function after autologous bone marrow-derived stem cell transfer in patients with acute myocardial infarction: a randomized, double-blind strain rate imaging study. *European Heart Journal*, vol. 30, 662 - 670 (2009)
10. McMahon, E. M., Korinek, J., Yoshifukub, S., Sengupta, P. P., Manducab, A., Belohlaveka, M. : Classification of acute myocardial ischemia by artificial neural network using echocardiographic strain waveforms. *Computers in Biology and Medicine*, vol. 38, 416 - 424 (2008)
11. Mitani, Y., Hamamoto, Y.: A local mean-based nonparametric classifier. *Pattern Recognition Letters*, vol 27, no. 10, 1151 - 1159 (2006)
12. Wold, S., Esbensen, K., Geladi, P.: *Principal Component Analysis*. *Chemometrics and Intelligent Laboratory Systems*, vol. 2, 37 - 52 (1987)